Available online at www.sciencedirect.com
SCIENCE DIRECT®
Ecological Modelling 191 (2006) 5–18
ECOLOGICAL MODELLING
www.elsevier.com/locate/ecolmodel

A new synergetic paradigm in environmental numerical modeling: Hybrid models combining deterministic and machine learning components

Vladimir M. Krasnopolsky [a,b,*], Michael S. Fox-Rabinovitz [a]

[a] *Earth System Science Interdisciplinary Center, University of Maryland, College Park, MD 20742, USA*
[b] *Science Application International Corporation at the National Centers for Environmental Predictions, NOAA, 5200 Auth Road, Camp Springs, MD 20746, USA*

Available online 10 October 2005

Abstract

A new type of environmental numerical models, hybrid environmental numerical models (HEMs) based on combining deterministic modeling and machine learning components, is introduced and formulated. Conceptual and practical possibilities of developing HEM, as an optimal synergetic combination of the traditional deterministic/first principles modeling (like that used for solving PDEs on the sphere representing model dynamics of global climate models) and machine learning components (like accurate and fast neural network emulations of model physics or chemistry processes), are discussed. Examples of developed HEMs (hybrid climate models and a hybrid wind–wave ocean model) illustrate the feasibility and efficiency of the new approach for modeling extremely complex multidimensional systems.

Keywords: Environmental modeling; Neural networks; Machine learning; Complex systems; Climate modeling

1. Introduction

Scientific and practical significance of interdisciplinary environmental numerical models (ENMs) increased tremendously during the last decades due to developments in numerical modeling and computing capabilities. Traditional ENMs are deterministic models based on first principle equations. For example, numerical atmospheric and oceanic models for climate and weather predictions are based on solving time-dependent three-dimensional geophysical fluid dynamics equations on the sphere. Physical and/or chemical processes (like radiation, convection, clouds, turbulence, chemical reactions) are so complicated that they can be included into ENMs only as one-dimensional (in the vertical direction) simplified or parameterized versions, as r.h.s. forcing for dynamics equations. Still, these parameterizations are the most time consuming components of ENMs. These parameterizations are usually derived using physical process

* Corresponding author. Tel.: +1 301 763 8000x7262; fax: +1 301 763 8545.
E-mail addresses: Vladimir.Krasnopolsky@noaa.gov, vladimir.krasnopolsky@ncep.noaa (V.M. Krasnopolsky), foxrab@essic.umd.edu (M.S. Fox-Rabinovitz).

0304-3800/$ – see front matter © 2005 Elsevier B.V. All rights reserved.
doi:10.1016/j.ecolmodel.2005.08.009

models and observational data and usually include some statistically derived empirical coefficients and some secondary empirical components based on traditional statistical techniques like regression. However, for this kind of ENMs, all major model components are deterministic, namely, not only model dynamics but also model physics and chemistry. Only recently the attempts were made to introduce statistical major components into ENMs. For example, an attempt was made to apply a traditional statistical technique, an expansion in hierarchical correlated functions, to approximate an atmospheric chemistry component (see Schoendorf et al., 2003 and the references therein). This traditional technique was successfully applied, however, with a limited accuracy. Much higher accuracy requirements must be met for such complex multidimensional systems as ENMs. A particular case of machine learning technique (MLT), namely, neural networks (NN), was successfully applied to speed up calculations of atmospheric and ocean model physics parameterizations (Chevallier et al., 1998, 2000; Krasnopolsky et al., 1999, 2002, 2004, 2005a,b; Krasnopolsky and Chevallier, 2001, 2003).

In this paper, on the basis of aforementioned preliminary studies and our current work with atmospheric and ocean physics, we formulate a new approach, which is based on a massive synergetic integration of deterministic and machine learning components in hybrid (combining deterministic modeling and machine learning) environmental models (HEM). We discuss conceptual and practical possibilities of developing HEMs, which are based on an efficient synergy or on an optimal combination of the traditional deterministic modeling like that of general circulation models or global climate models (GCMs) and MLTs for accurate and fast emulation of model physics. In Section 2, we describe HEM using accurate and fast emulations based on MLT to solve a major computational problem in development of high-quality high-resolution environmental numerical models. In Section 3, we formulate our approach. In Section 4, we present several developed HEMs. Section 5 contains conclusions and discussion.

2. Hybrid models

One of the main problems of development and implementation of high-quality high-resolution environmental models is the complexity of physical, chemical and biological processes involved. We will discuss MLT emulation for model physics keeping in mind that the discussion is applicable to model chemical and biological processes as well. Parameterizations of model physics (chemistry, etc.) are approximate, adjusted to model resolution and computer resources, schemes based on simplified physical process equations and empirical data. Still, the parameterizations are so time-consuming, even for most powerful modern supercomputers, that some of them have to be calculated less frequently than model dynamics. This may negatively affect the accuracy of ecological simulations and predictions. For example, in the case of a very sophisticated ENM—climate model, calculation of a model physics package in a typical moderate (a few degrees) resolution GCM like the National Center for Atmospheric Research (NCAR) Community Atmospheric Model (CAM) takes about 70% of the total model computations. Higher uniform and variable model resolutions (e.g. Fox-Rabinovitz et al., 2002; Duffy et al., 2003) and more frequent model physics calculations, desirable for temporal consistency with model dynamics, would increase the computation time for model physics.

Such a situation is an important motivation for looking for alternative, faster, and most importantly, very accurate ways of calculating model physics, chemistry and biology. The approach discussed in this paper introduces a new paradigm of hybrid numerical model, which is based on a synergetic combination of deterministic numerical modeling with machine learning techniques for emulating model physics.

During the last decade a new emerging approach based on machine learning NN approximations has found the variety of applications in different fields and, more specifically, for accurate and fast approximation of atmospheric radiative processes (Krasnopolsky, 1997; Chevallier et al., 1998), and for environmental satellite data processing (Stogryn et al., 1994; Krasnopolsky et al., 1995; Krasnopolsky and Schiller, 2003). Recently, the NN approach has been successfully applied for developing a fast (eight times faster than the original parameterization) and accurate long-wave (LW) radiation parameterization for the European Centre for Medium-range Weather Forecasting (ECMWF) model (Chevallier et al., 2000). This LW radiation parameterization is being

used operationally within the ECMWF 4-dimensional variational data assimilation system since October 2003. The NN emulation approach has been also used for emulations of model physics in ocean and atmospheric numerical models (Krasnopolsky et al., 1999, 2002, 2004, 2005a,b; Tolman et al., 2005) where acceleration of calculation from 10 to 10^5 times has been achieved, as compared with original parameterizations (see Section 4).

Based on above results we introduce a new notion of hybrid ENM (HEM), which combines deterministic (e.g. model dynamics) and machine learning (e.g. NN emulations of model physics and/or chemistry) components to perform calculations more effectively than original completely deterministic ENM. To evaluate feasibility of HEM, four key questions of the new approach should be answered: (i) are developed machine learning approximations close enough to the original physical/chemical parameterizations so that their use (instead of the original parameterization) allows us to preserve all richness and complete integrity and all the detailed features of environmental physical/chemical processes, (ii) are these approximations fast enough to significantly accelerate calculations of model physics/chemistry, (iii) are these statistical/machine learning techniques able to successfully coexist with deterministic components of ENMs so that their combination (which is HEM) can be efficiently used for accurate and fast environmental simulations without any negative impacts on their quality, and (iv) is there a real/productive synergy here, in other words, does this new combination of deterministic and statistical learning approaches, the new HEM, lead to new opportunities in environmental simulations.

3. Formulation of the hybrid approach

3.1. Hybrid approach

A NN *emulation* of a model physics parameterization is a functional imitation of this parameterization so that the results of model calculations with the original parameterization and with its NN emulation are physically (and climatologically) identical. High quality of NN emulations is achieved due to the high accuracy of approximation of the original components. We prefer to use the term NN emulation but not NN approximation to avoid a possible confusion. The term parameterization already means a simplified approximation of physical processes. So, in the context of our approach, the term emulation means a complete functional imitation based on a precise mathematical/statistical approximation (in a classic mathematical sense) of model physics parameterizations.

The key point is that NN emulation is developed here for the *existing* parameterizations of model physics. This allows us to preserve the integrity and the level of sophistication of the state-of-the-art parameterizations of physical processes. Due to the capability of modern MLTs to provide an unprecedented accuracy for approximation of complex systems like model physics, our NN emulations of model physics parameterizations are practically identical to original physical parameterizations. As a result, HEM using this emulation produces results, which are physically identical to those of the original ENM. In other words, the underlying idea of the approach is not developing a new parameterization but rather emulating a parameterization already very carefully tested and validated by its developers off-line and then on-line through experimentation with the entire model. It is achieved by using for NN training data simulated by running an original model (i.e. ENM) with the original parameterization. Using model-simulated data for NN training allows us to achieve a very high accuracy for approximation because simulated data are free of the problems typical for empirical data (problems like high level of observational noise, sparse spatial and temporal coverage, poor representation of extreme events, etc.). In the context of our approach, the accuracy and improved computational performance of HEM and NN emulations is always measured against the ENM using the original parameterization. It is noteworthy that the developed NN emulation has the same inputs and outputs as the original parameterization and is used as its functional substitute in the model.

3.2. Neural network technique

NN emulations of model physics/chemistry are based on the fact that any parameterization of model physics/chemistry can be considered as a continuous or almost continuous mapping (output vector versus input vector dependence), and NNs (multilayer perceptrons in our case) are a generic tool for approximation of such

mappings. NN is an analytical approximation that uses a family of mappings like

$$y_q = a_{q0} + \sum_{j=1}^{k} a_{qj}\phi\left(b_{j0} + \sum_{i=1}^{n} b_{ji}x_i\right);$$
$$q = 1, 2, \ldots, m \quad (1)$$

where x_i and y_q are the components of the input and output vectors respectively, a and b the fitting parameters, ϕ a so-called activation function (usually it is a hyperbolic tangent), n and m the numbers of inputs and outputs, respectively, and k is the number of neurons in the hidden layer (see Ripley, 1997; Appendix in Krasnopolsky et al., 2002 for more details).

3.3. Development and validation framework

Let us formulate a developmental framework and validation criteria that are, in our view, useful, instrumental, and recommended to be followed when developing and validating machine learning components of HEM, NN emulations of model physics (or chemistry) components. The developmental process consists of *three major steps*.

The *first step* is the problem analysis or the analysis of the model component to be approximated (for example, the original parameterization). The purpose of this analysis is first to determine the topology (architecture) of the future NN emulation by specifying the number of NNs to be developed. There are several ways to develop NN emulations for model physics components. Namely, the physical or computational structures of a model physics scheme can be analyzed and followed. In the first example of the next section, the physical structure of the original parameterization has been followed that resulted in developing a single NN emulating this parameterization. The second example, the nonlinear wave–wave interaction, demonstrates another case when the straightforward approach used in the first case does not work and a more sophisticated approach should be implemented. Yet another way would be following the computational structure of a parameterization, determining the computational "bottlenecks" and developing NN emulation for each of them. The second purpose of the topological analysis is to determine all the inputs and outputs, and their number for each particular NN. For any structuring strategy, the number of inputs and outputs for each NN should not exceed a reasonable necessary limit. The analysis of the complexity of the original parameterization will help in selecting the initial number of neurons in the hidden layer.

The *second step* is generation of representative data sets for training, testing and validation. The number of records and the time interval covered by these data sets depend on such characteristics of a model as resolution, a type of a model (climate, weather, or chemistry model), and on a particular type of the model component under consideration. This approach is based on using simulated data, which allows us to produce NN emulations physically identical to the original parameterizations. To create a representative data set the original ENM is run long enough to produce all possible atmospheric states, phenomena, etc. To account for insufficient sampling for some events it is possible to run the original parameterization off-line generating complimentary data to extend sampling. In some applications, the use of blended (simulated, assimilated, and observational) data for NN training could be beneficial.

The *third step* is the NN training. Several different versions of NNs with different number of neurons in one (or several) hidden layers should be trained to determine the optimal size of the hidden layer, which provides both sufficient approximation accuracy and reasonable acceleration of calculations. For each of these NNs, several initialization procedures and training algorithms should be applied to assure that good minimum of the error function is found.

Validation of HEM with trained NN emulation consists of two major steps. The first step is validation of NN approximation against the original parameterization using the independent validation data set. Both the original parameterization and its NN emulation are complicated multidimensional objects (mappings). Many different statistical metrics of the approximation accuracy should be calculated to assure that the sufficiently complete evaluation of the approximation accuracy is obtained. For example, total, level, and profile statistics have to be evaluated (see Section 4). The second validation step consists of comprehensive analysis of parallel HEM versus ENM runs. For the parallel experiments, all relevant model prognostic and diagnostic fields should be analyzed and carefully compared to assure that the integrity of ENM and the original parameterization, with all its details and

characteristic features, is preserved when using HEM with NN emulation (see Section 4.1). Therefore, the development and application framework of the new synergetic approach should be focused on obtaining the high accuracy for both NN emulation and HEM simulations.

4. Examples of machine learning components and HEMs

4.1. Atmospheric applications

In this section we present applications of the approach formulated in the previous sections to atmospheric models. Two atmospheric models, NCAR CAM (the two latest versions of this model have been used for our developments and experiments) and NASA's NSIPP (NASA Seasonal-to-Interannual Prediction Program) GCM, are used here as examples of complex ENMs that include parameterizations of model physics. NCAR GCM is the state-of-the-art widely recognized model used by a large community for long term climate prediction. NSIPP GCM was extensively used for both short and long term climate prediction. More specifically, in this Section we apply the NN and HEM methodology and framework to the NCAR CAM radiation block including the long-wave radiation (LWR) and the short-wave radiation (SWR) parameterizations, and to the LWR parameterization in NSIPP GCM. NCAR CAM and NSIPP GCM are complex climate models with different model dynamics and physics. They have different model dynamics, vertical and horizontal resolutions and different formulations for atmospheric physics including LWR parameterizations. NCAR CAM uses a spectral discetization for model dynamics, which has 42 spectral components (corresponds approximately to about 3.5° horizontal global resolution) and 26 vertical levels. The NSIPP model used in this study is a grid point model, which has horizontal (latitude × longitudes) resolution 2° × 2.5° and 40 vertical levels. Applying the NN approach to these different models is aimed at investigating the robustness of the approach.

4.1.1. Statistical metrics for validation of NN emulations

Our NN emulations have been validated against the corresponding original model, NCAR CAM or NSIPP, parameterization. Mean difference, B (bias or a systematic error of approximation), and the root mean square difference (a root mean square error of approximation), RMSE, between the original parameterization and its NN emulation, are calculated as follows:

$$B = \frac{1}{NL}\sum_{i=1}^{N}\sum_{j=1}^{L}[Y(i,\,j) - Y_{\mathrm{NN}}(i,\,j)],$$

$$\mathrm{RMSE} = \sqrt{\frac{\sum_{i=1}^{N}\sum_{j=1}^{L}[Y(i,\,j) - Y_{\mathrm{NN}}(i,\,j)]^2}{NL}} \qquad (2)$$

where $Y(i,j)$ and $Y_{\mathrm{NN}}(i,j)$ are the outputs of the original parameterization and of NN emulation, respectively, where i = (lat, lon), $i = 1, \ldots, N$ the horizontal location of a vertical profile, N the number of horizontal grid points, and $j = 1, \ldots, L$ is the vertical index where L is the number of the vertical levels. Using a slight modification of Eqs. (2), bias B_m and RMSE_m for the mth vertical level can be calculated as follows:

$$B_m = \frac{1}{N}\sum_{i=1}^{N}[Y(i,m) - Y_{\mathrm{NN}}(i,m)],$$

$$\mathrm{RMSE}_m = \sqrt{\frac{\sum_{i=1}^{N}[Y(i,m) - Y_{\mathrm{NN}}(i,m)]^2}{N}} \qquad (3)$$

4.1.2. NN emulations for the NCAR CAM LWR and SWR and for NSIPP LWR parameterizations

The functions of the LWR and SWR parameterizations in atmospheric GCMs are to calculate the corresponding heating/cooling fluxes and rates. The complete description of the NCAR CAM atmospheric LWR and SWR parameterizations is presented in (Journal of Climate, 1998; Collins, 2001). The NASA NSIPP LWR parameterization is described in Chou et al. (2001). The atmospheric radiation is the most time consuming part of the atmospheric model physics. For NCAR CAM the calculation of the radiation block (LWR and SWR) takes about 70% of time required for calculating the model physics and about 50% of the total calculation time. This consideration was our main motivation for selecting the atmospheric radiation block as the first candidate for applying our NN emulation and HEM approach.

Because our NN emulations are functional imitations of the original parameterizations, they have the

same inputs and outputs as the original parameterizations, which are determined by the internal structure of physical equations and approximations implemented within the original parameterizations. Because of that the NCAR CAM and NSIPP LWR parameterizations and their corresponding NN emulations have slightly different inputs and outputs as the result of the two different physical schemes implemented in the original parameterizations (compare Collins, 2001; Chou et al., 2001).

The input vectors for the NCAR CAM LWR parameterization and our corresponding NN emulation contain totally 220 inputs ($n = 220$ in Eq. (1)). The inputs include 10 profiles: atmospheric temperature, humidity, ozone, CO_2, N_2O, CH_4, two CFC mixing ratios (the annual mean atmospheric mole fractions for the halocarbons), pressure, and cloudiness and one relevant surface characteristic (upward LW flux at the surface). The total number of outputs is 33 ($m = 33$ in Eq. (1)). The output vectors consist of the profile of heating rates (HRs) and several radiation fluxes including the outgoing LWR flux from the top layer of the model atmosphere (the outgoing LW radiation or OLR).

The input vector for the NSIPP LWR includes five vertical profiles (cloud fraction, pressure, temperature, specific humidity, and ozone mixing rate) and the surface temperature, or totally 202 inputs. The NSIPP LWR output vector consists of the profile of heating rates and one surface parameter, or the total of 41 outputs (Krasnopolsky et al., 2005b).

The input vectors for the NCAR CAM SWR parameterization include 21 profiles (specific humidity, ozone concentration, pressure, cloudiness, aerosol mass mixing ratios, etc.) and several relevant surface characteristics. The CAM SWR parameterization output vectors consist of the profile of heating rates (HRs) and several radiation fluxes. This parameterization and our corresponding NN emulation have for each the total of 451 inputs and 33 outputs.

For each of these parameterizations we developed several NNs, all of which have one hidden layer, with 100, 150, and 200 neurons ($k = 100, 150, 200$ in Eq. (1)). Varying the number of hidden neurons allows us to demonstrate the dependence of the accuracy of approximation and its convergence on this parameter, and as a result, to provide the sufficient accuracy of approximation for the climate models (for a detailed discussion of conversions of the approximations, see Krasnopolsky et al., 2005a).

The ENMs, NCAR CAM and NSIPP, were run for 2 years to generate the representative data sets. The first year of model simulation was divided into two independent parts each containing input/output vector combinations. The first part was used for training and the second one was used for tests (control of overfitting, control of a NN architecture, etc.). The second year of simulation was used to create a validation data set completely independent from both training and test ones. This third part was used for validations only. All approximation statistics presented below in this section are calculated using this independent validation data set. The data sets simulated by NCAR CAM were used for training NN emulations for the CAM radiation (LWR and SWR) block and data simulated by NSIPP model for training NN emulations for the NSIPP LWR.

Three different NN emulations with different number of neurons (100, 150, and 200) in one hidden layer have been developed for each parameterization. All these NN emulations have almost zero or negligible systematic errors (biases) (see Table 1), which practically do not depend on height. Obtaining such small biases is important for preventing an undesirable error accumulation during long term model calculations discussed below. Table 1 demonstrates the conversions of NN emulations in terms of approximation errors when the number of hidden neurons increases. Table 1 and Fig. 1 also demonstrate the robustness of our NN emulation technique, which produces the approximation errors of practically the same magnitude when applied to three different parameterizations in two very different models. Fig. 1 shows the vertical profiles of RMSEs (2), which in the case of zero biases (or practically zero biases as in our case) are purely random errors of approximation, for six NN emulations developed for the CAM radiation block (LWR and SWR). Three NNs with different number of neurons (100, 150, and 200) in one hidden layer have been developed for each LWR and SWR parameterizations. For all NNs, the RMSE for the 10 upper levels does not exceed 0.2 K/day reaching 0.4 K/day at the 23rd level. For the two lowest levels, RMSE is about 0.6–0.8 K/day for the LWR and only 0.4–0.5 K/day for SWR. Statistics (biases and RMSEs) for the lowest (26th) level are also included in Table 1. The natural variability (σ) of the HRs is significantly higher (see in the title of Table 1) at

Table 1
Statistics estimating the accuracy of HRs (in K/day) calculations and computational performance for NCAR CAM LWR and SWR and NSIPP LWR using NN emulation vs. the original parameterization

NN size	Model/parameteriazation	Bias	RMSE	Bias_{ll}	RMSE_{ll}	Performance
NN100	CAM/LWR	-9×10^{-4}	0.30	3×10^{-3}	0.77	∼80 times faster
	CAM/SWR	-6×10^{-3}	0.22	-9×10^{-3}	0.48	
	NSIPP/LWR	-2×10^{-4}	0.58	7×10^{-3}	0.97	
NN150	CAM/LWR	9×10^{-4}	0.25	1×10^{-2}	0.67	∼50 times faster
	CAM/SWR	-4×10^{-5}	0.20	2×10^{-2}	0.47	
	NSIPP/LWR	-6×10^{-4}	0.47	1×10^{-3}	0.79	
NN200	CAM/LWR	1×10^{-3}	0.21	-1×10^{-2}	0.56	∼35 times faster
	CAM/SWR	-5×10^{-3}	0.18	-2×10^{-2}	0.40	
	NSIPP/LWR	-2×10^{-4}	0.45	3×10^{-3}	0.76	

Bias_{ll} and RMSE_{ll} (in K/day) correspond to the lowest layer (26th for CAM and 40th for NSIPP). The total mean value for HRs = −1.36 K/day and the standard deviation $\sigma_{\text{HR}} = 1.93$ K/day. For the lowest CAM level (26th), the mean value for HRs = −2.22 K/day and $\sigma_{\text{HR}} = 5.57$ K/day. NN*nnn* denotes NN with *nnn* neurons in the hidden layer.

lowest levels than at higher ones. Hence, relative errors in the HRs calculated with respect to the natural variability (σ) are approximately the same as errors for the higher levels.

Table 1 shows a bulk validation statistics for the accuracy of approximation and computational performance for the three best (in terms of their accuracy and performance) developed NN emulations. Mean values and standard deviations (σ_{HR}) of HRs are presented in the title of Table 1 for a better understanding of relative errors. Therefore, the obtained RMSEs are sufficiently small for the entire vertical profile compared to the mean and standard deviation values.

In addition to this high approximation accuracy, our NN emulation performs about 80–35 times faster (for NN100, NN150, and NN200, respectively) than the original parameterizations. Table 1 and Fig. 1 clearly show a systematic improvement of the accuracy of approximation with the increase of the size of the NN hidden layer. Table 1 also demonstrates a corresponding reduction, when increasing the number of neurons, of the performance gain from 80 to 35 times faster than the original parameterization. This offers an opportunity for the accuracy versus performance trade off. However, as we mentioned earlier, in this trade off the key requirement, which allows us to obtain the successful, synergetic functioning of the NN emulations within HEMs, is to preserve the accuracy and integrity of the original parameterization, i.e. the detailed description of the corresponding physical process. Obviously, the final decision on the optimal NN version, which has to be implemented into the model, should be made based on testing these NNs in HEM simulations.

Both the original parameterization and its NN emulation are complicated multidimensional objects (mappings). In this case, calculating bulk statistics is not sufficient for evaluating the accuracy of approximation. We evaluated many different statistical metrics of the approximation accuracy (Krasnopolsky et al., 2005a); some of them are shown in Figs. 1 and 2.

Fig. 2 shows the absolute zonal mean bias (the left column) and zonal mean RMSE (the right column). The zonal mean error is a three-dimensional error field (error as a function of latitude, longitude, and height) averaged over the longitude and presented as a two-dimensional field in latitude–height coordinates. Three panel rows (top, middle, and bottom) correspond to three different NNs, namely NN100, NN150, and NN200, respectively. With increasing the accuracy of approximation (when increasing the number of hidden neurons in the NN) both zonal mean bias and RMSE decrease significantly. Comparing the top, middle and bottom panels, we see that small spots with the bias > 0.01 K/day (the left column) in the lower part of the atmosphere disappear completely. Also, the small spots of RMSE > 0.25 K/day (the right column) disappear at the upper levels. In the lower part of the atmosphere, small areas with RMSE > 1 K/day (the right column) disappear already for NN150, and even more so for NN200, and the areas with RMSE > 0.5 K/day are confined just within two small spots located in the polar areas.

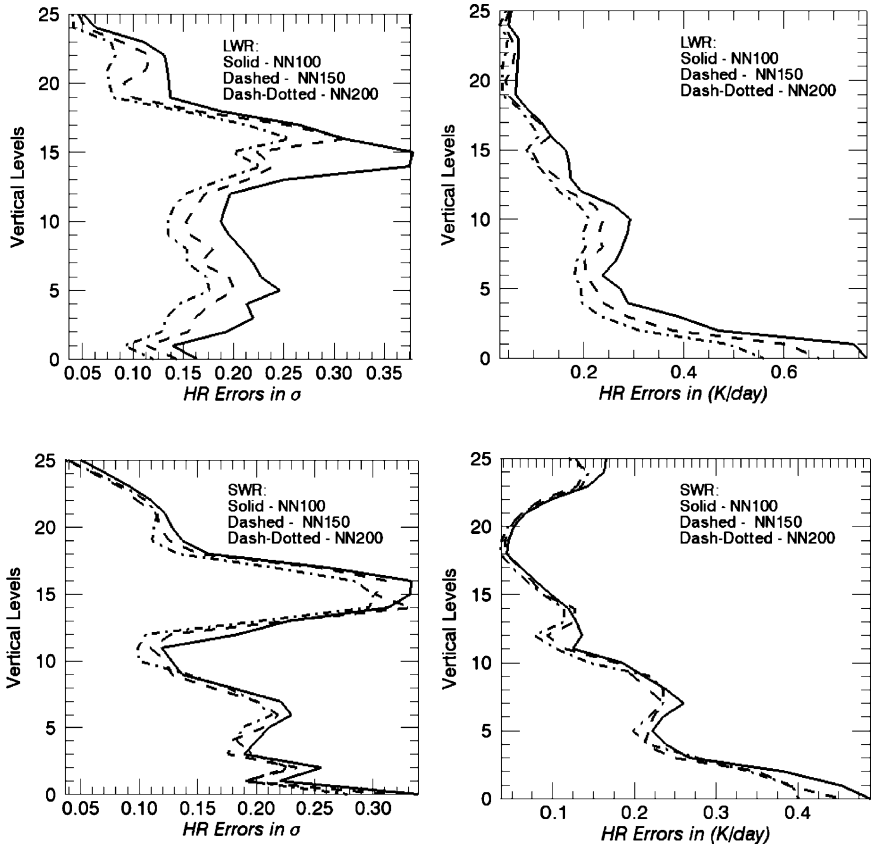

Fig. 1. The vertical profiles of RMS approximation errors (3) for NN emulations of LWR (upper row) and SWR (lower row) at each of 26 levels of CAM. Left panels: RMSEs in the units of the standard deviations of HR for corresponding vertical level. Right panels: RMSEs in absolute units (K/day). Three NNs are shown for each parameterization—NN100: solid, NN150: dashed, and NN200: dash-dotted.

4.1.3. Validation of LWR NN emulations in a long term climate model simulation

The analysis of approximation errors presented above shows that the NN technique is capable of providing NN emulations with practically zero systematic errors or biases and small random errors. The next step is validation of NN emulations in a long term climate model simulation. For assessing the impact of using NN emulation of the LWR parameterization in the HEM, the parallel climate simulation runs have been performed with the original ENM, NCAR CAM including the original LW radiation parameterization (the control run), and with the HEM, NCAR CAM including our NN emulations described above. The climate simulations have been run for 10 years started after the training and validation period, namely for years 3 through 12. All the comparisons of the control and NN emulation runs presented below in the section are done by analyzing the time (10-year) mean differences between the results of different runs.

Mass preservation is the most important property for climate simulations. In the climate simulations performed with the original ENM and with HEM the time mean global mass or mean surface pressure is precisely preserved. For example, for the NN150 run there is a negligible difference of 0.0001% between the NN and control runs (see Table 2). Other time global means, some of which are also presented in Table 2, show a profound similarity between the simulations in these terms, with the differences usually within about 0.03–0.1% and not exceeding 0.1–0.3%. Other simulations (with NN100 and NN200) show similar results.

We also carefully investigated key simulated climate diagnostic and prognostic fields and their differences

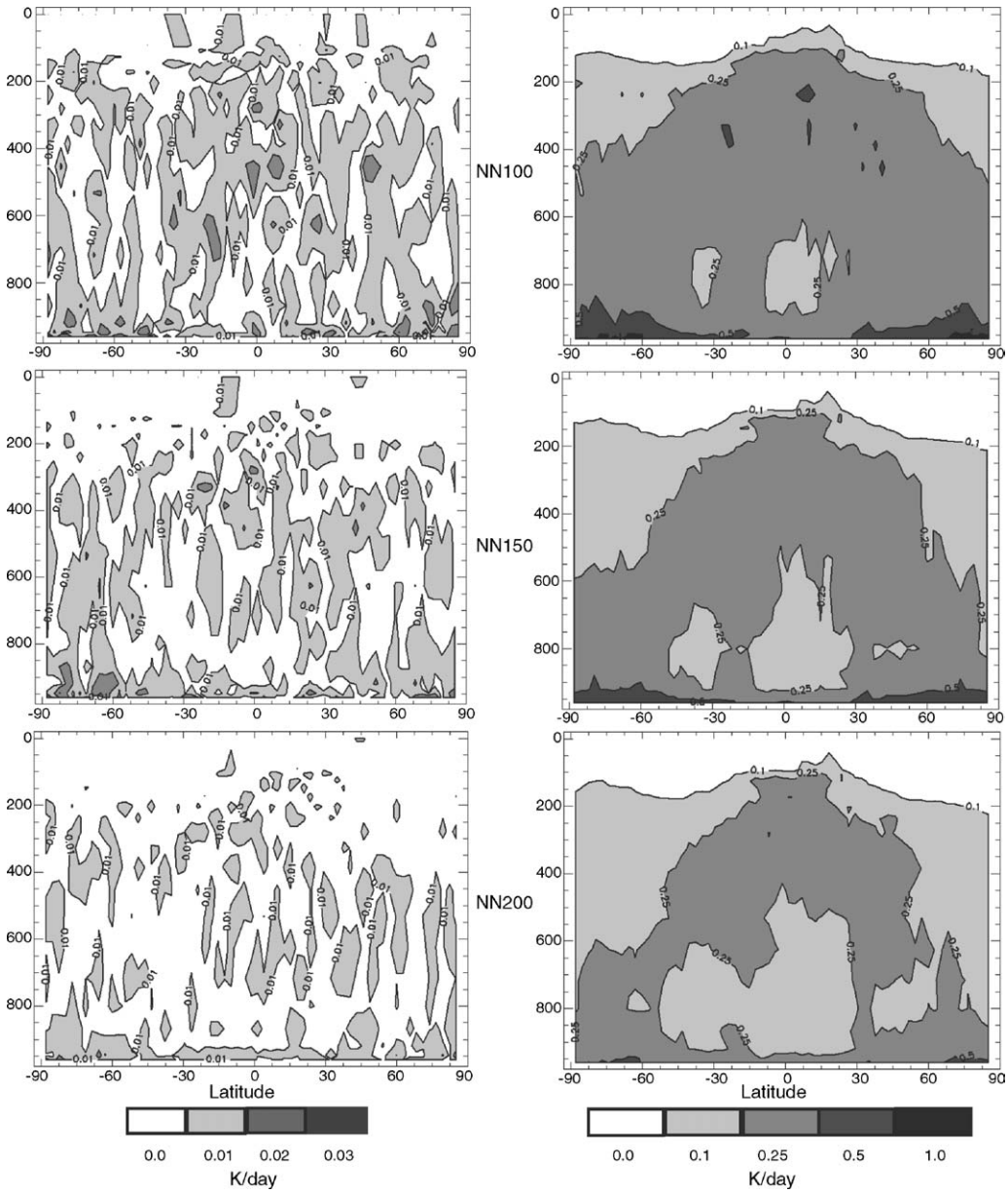

Fig. 2. Absolute zonal mean bias (the left column) and zonal mean RMSE (the right column) for NN100, NN150, and NN200 (for the top, middle and bottom panels, respectively).

produced in the parallel control (original ENM) and NN emulation (HEM) runs. The comparison of biases for NN100 and NN150 runs (Krasnopolsky et al., 2005a) confirms that increasing the number of hidden neurons from 100 to 150 leads to a measurable bias reduction that positively affects the accuracy of the NN150 climate simulation in terms of its profound similarity to the control simulation. Most importantly, biases for both NN100 and NN150 10-year simulations are not accumulating in time. It is noteworthy that even the maximum temperature, zonal wind and moisture biases for the NN100 run are still below the corresponding

Table 2
Time and global means for mass (mean sea level pressure) and other model diagnostics for the NCAR CAM climate simulations with the original LWR parameterization and its NN emulation (NN150) and their differences (in %)

Field	Original LWR parameterization	NN emulation	Difference (%)
Mean sea level pressure (hPa)	1011.480	1011.481	0.0001
Surface temperature (K)	289.003	289.001	0.0007
Total precipitation (mm/day)	2.275	2.273	0.09
Total cloudiness (fractions 0.1–1)	0.607	0.609	0.3
LWR heating rates (K/day)	−1.698	−1.700	0.1
Outgoing LWR–OLR (W/m^2)	234.43	234.63	0.08
Latent heat flux (W/m^2)	82.84	82.82	0.03

observational errors. These results are practically significant and constitute the proof of the concept for HEM.

4.2. Ocean application: a neural network approximation for nonlinear interactions in wind wave models

Another ENM, ocean wind wave model for simulation and forecast purposes, is based on a form of the spectral energy or action balance equation:

$$\frac{DF}{Dt} = S_{in} + S_{nl} + S_{ds} + S_{sw} \tag{4}$$

where F is the spectrum, S_{in} the input source term, S_{nl} the nonlinear wave–wave interaction source term, S_{ds} the dissipation or 'whitecapping' source term, and S_{sw} represents the additional shallow water source terms. State-of-the-art wave models explicitly model S_{nl} source term.

In its full form (e.g. Hasselmann and Hasselmann, 1985), the calculation of the interactions S_{nl} requires the integration of a six-dimensional Boltzmann integral:

$$\begin{aligned} S_{nl}(\vec{k}_4) &= T \otimes F(\vec{k}) \\ &= \omega_4 \int G(\vec{k}_1, \vec{k}_2, \vec{k}_3, \vec{k}_4)\delta(\vec{k}_1 + \vec{k}_2 - \vec{k}_3 - \vec{k}_4) \\ &\quad \times \delta(\omega_1 + \omega_2 - \omega_3 - \omega_4) \\ &\quad \times [n_1 n_3 (n_4 - n_2) + n_2 n_4 (n_3 - n_1)] \\ &\quad \times d\vec{k}_1\, d\vec{k}_2\, d\vec{k}_3, \quad n(\vec{k}) = \frac{F(\vec{k})}{\omega}; \\ \omega^2 &= gk \tanh(kh) \end{aligned} \tag{5}$$

where the complicated coupling coefficient G contains moving singularities. This integration requires roughly 10^3–10^4 times more computational effort than all other aspects of the wave model combined. Present operational constraints require that the computational effort for the estimation of S_{nl} should be of the same order of magnitude as the remainder of the wave model. This requirement was met with the development of the discrete interaction approximation (DIA, Hasselmann et al., 1985). More than two decades of experience with the DIA in wave models has identified significant shortcomings of the DIA.

Considering the above, it is of crucial importance for the development of third generation wave models to develop *an economical yet accurate approximation* for S_{nl}. We explored a neural network interaction approximation (NNIA) to achieve this goal (see also Krasnopolsky et al., 2002; Tolman et al., 2005). NNs can be applied here because the nonlinear interaction (5) is essentially a nonlinear mapping (symbolically represented in Eq. (5) by T) which relates two vectors, F and S_{nl} (two-dimensional fields in this case). Discretization of S and F (as is necessary in any numerical approach) reduces (5) to continuous mapping of two vectors of finite dimensions. Modern high resolution wind wave models use descretization on a two-dimensional grid which leads to dimensions of S and F vectors of order of $N \sim 1000$. It seems unreasonable to develop a NN approximation of such a high dimensionality (more than 1000 inputs and outputs). Moreover, such a NN will be grid dependent.

In order to reduce the dimensionality of the NN and convert the mapping (5) to a continuous mapping of two finite vectors less dependent on the actual spectral discretization, the spectrum F and source function

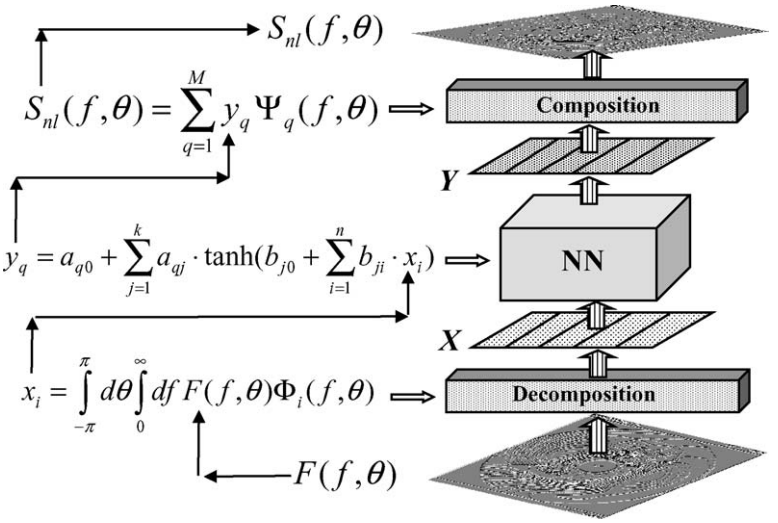

Fig. 3. Graphical representation of the NNIA and NNIAE algorithms.

S_{nl} are expanded using systems of two-dimensional functions each of which ($\boldsymbol{\Phi}_i$ and $\boldsymbol{\Psi}_q$) creates a complete and orthogonal two-dimensional basis:

$$F \approx \sum_{i=1}^{n} x_i \boldsymbol{\Phi}_i, \qquad S_{\mathrm{nl}} \approx \sum_{q=1}^{m} y_q \boldsymbol{\Psi}_q \tag{6}$$

where for coefficients of decomposition/composition x_i and y_q, we have

$$x_i = \iint F \boldsymbol{\Phi}_i, \qquad y_q = \iint S_{\mathrm{nl}} \boldsymbol{\Psi}_q, \tag{7}$$

where the double integral identifies integration over the spectral space. The developed NN relates vectors of coefficients $\boldsymbol{X}$ and $\boldsymbol{Y}$: $\boldsymbol{Y} = T_{\mathrm{NN}}(\boldsymbol{X})$.

To train the NN approximation, T_{NN}, a training set has to be created that consists of pairs of vectors $\boldsymbol{X}$ and $\boldsymbol{Y}$. To create this training set, a representative set of spectra F_p has to be generated with corresponding (exact) interactions $S_{\mathrm{nl},p}$ using Eq. (5). For each pair $(F, S_{\mathrm{nl}})_p$, the corresponding vectors $(\boldsymbol{X}, \boldsymbol{Y})_p$ are determined using Eq. (7). These pairs of vectors are then used to train the NN to obtain T_{NN}. After T_{NN} has been obtained by training, the resulting NN interaction approximation (NNIA) algorithm consists of three steps: (i) decompose the input spectrum, F, by applying Eq. (7) to calculate $\boldsymbol{X}$; (ii) estimate $\boldsymbol{Y}$ from $\boldsymbol{X}$ using NN; (iii) compose the output source function, S_{nl}, from $\boldsymbol{Y}$ using Eq. (6). A graphical representation of the NNIA algorithm is shown in Fig. 3.

Two approaches have been used for the basis functions. The first is a mathematical basis used in (Krasnopolsky et al., 2002). As it is usually done in the parametric spectral description of wind waves, separable basis functions are chosen where the frequency and angular dependence are separated. The advantage of this choice of basis functions is the simplicity of the basis generation. The disadvantage is the slow convergence of the decompositions. As an alternative, a second approach to the basis functions choice has been investigated. In this approach, empirical orthogonal functions (EOFs) or principal components (Lorenz, 1956; Jolliffe, 2002) are used (Tolman et al., 2005).

EOFs form a statistically optimal basis. In the present case, the basis functions $\boldsymbol{\Phi}_i$ and $\boldsymbol{\Psi}_q$ are functions of two variables f and θ. The set of spectra F and source terms S_{nl}, which are used for the training of the NN, are also used to generate the EOFs for decomposing F and S_{nl}. When using EOFs, the basis generation procedure is computationally expensive, with the cost increasing as the resolution of the model increases. However, like the NN training, the basis generation needs to be performed only once. Stored results can be used without the need for re-calculation in a practical

Table 3
Approximation RMSEs (in nondimensional units) and performance (see units in text) for DIA, NNIA, NNIAE, and exact S_{nl} calculation (original)

Algorithm	RMSE	Performance
DIA	0.312	1
NNIA	0.088	4
NNIAE	0.035	7
Original parameterization	0	$\sim 8. \times 10^5$

NNIA. The main advantage of EOFs is the fast convergence of the decomposition.

To distinguish between NN algorithms using different bases functions for decomposition, we use abbreviation NNIAE for our NN algorithm, which used the EOF basis. Table 3 demonstrates comparisons of the accuracy and performance of DIA with those of two NN emulations (NNIA and NNIAE) all versus the exact calculation of S_{nl} (original parameterization). Approximation errors (RMSEs) are calculated in nondimensional units and performance is measured in DIA calculation times (taken as a unit). The NNIAE is nearly 10 times more accurate than DIA. It is about 10^5 times faster than the original parameterization. As for the case of the atmospheric long wave radiation a careful investigation of parallel runs of original ENM (wave model with original wave–wave interaction) and HEM with NN emulation is being performed currently.

5. Conclusions and discussion

In the study, we formulated a new paradigm in the environmental numerical modeling. We introduced a new type of ENM—a hybrid environmental model (HEM) which is based on a synergetic combination of deterministic modeling and a machine learning technique within an HEM. This approach uses neural networks as a statistical or machine learning technique to develop highly accurate and fast approximations for model physics/chemistry components. The obtained results show:

(i) the conceptual and practical possibility of developing HEMs with accurate NN emulation of model components, which preserves the integrity and all the detailed features of original ENM;
(ii) that accurate NN emulations are robust and very fast (up to 10^5 times faster than original parameterization) so that the significant acceleration of HEM calculations can be achieved without compromising its accuracy;
(iii) that statistical/machine learning components can be successfully combined with deterministic model components within HEM so that their synergy can be efficiently used for environmental simulations without any negative impacts on simulation quality;
(iv) that this productive synergy or the new combination of the state-of-the-art deterministic and statistical learning approaches leads to new opportunities in HEM for environmental simulations and prediction. For example, new more sophisticated parameterizations and even "superparameterizations", that are computationally prohibited when used in their original form in ENM, will become computationally "affordable" when using their accurate and computationally more efficient MLT emulations in HEM.

This study presents the major framework and first experimental results for new hybrid modeling approach. Let us briefly outline possible/potential future avenues of the developments of the hybrid modeling framework.

The development of NN emulations, the core of the hybrid modeling approach, depends significantly on our ability to generate a representative training set to avoid using NN for extrapolation far beyond the domain covered by the training set. Because of high dimensionality of the input domain which is of the order of several hundreds or more, it is rather difficult to cover the entire domain, especially its "far corners" associated with rare events, even when we use the simulated data for the NN training. Another related problem is that NN emulations are developed for a changing in time environmental system. It means that for a climate simulation the domain configuration may change in time (for example, when using a future climate change scenario). For both described situations, the emulating NN may be forced to extrapolate beyond its generalization ability that may lead to errors in NN outputs and result in the corresponding HEM simulation errors.

To take care about this kind of problems and make our NN emulation approach suitable for long term cli-

mate change simulations and other applications, we are developing two new techniques: a *compound parameterization* (CP) and an NN *dynamical adjustment* (DA). Their full description will be provided on the completion of the developments. Here we will just briefly outline them.

The compound parameterization consists of three elements: the original parameterization, its NN emulation, and a quality control (QC) block. During a routine HEM simulation with CP, the NN emulation is used by default, and it generates physical parameters (outputs), which are quality controlled. If QC accepts the output parameters, they are used in HEM. If QC rejects the NN emulation outputs, the original parameterization is used instead for generating the physical parameters, which are used in HEM. When the original parameterization is used instead of the NN emulation, its inputs and outputs are saved for a further adjustment of the NN emulation. After accumulating a sufficient number of the records, a dynamical adjustment of the NN emulation is produced by a short retraining using the accumulated input/output records. Thus, the adapted NN emulation becomes dynamically adjusted to the changes and/or new events/states produced by the complex environmental system.

There are different possible designs, which we consider for QC. The first and simplest QC design is based on a set of regular physical and statistical tests that are used for checking the consistency of the NN outputs. The second QC design is based on training additional NNs specifically for estimating the errors in the NN emulation outputs. If these errors exceed a predefined threshold, the original parameterization is used instead of NN emulation. The third and most promising QC design is based on the domain check technique proposed in the context of NN applications to satellite remote sensing [8]. In this case, QC is based on a combination of forward and inverse NNs.

In this paper only one MLT, the NN technique, has been discussed and investigated for two types of HEMs considered. We started from NNs because this technique is well established and developed. However, it is not optimal. There are new techniques like support vector machines (SVM) and related approaches which may provide optimal approximation for obtaining MLT components with potentially even better accuracy and performance as compared to NNs. Also, other types of ENMs can be considered and faster HEMs developed using the approach described in this paper. We are planning on investigating other MLTs and other types of HEMs in the follow-up studies.

Acknowledgments

The authors would like to thank Drs. D. Chalikov and H. Tolman for useful discussions and fruitful collaboration. The authors would like to thank Drs. W. Collins, P. Rasch, J. Tribbia and S. Lord for helpful consultations and discussions. The authors acknowledge the programming support provided by Mr. R. Govindaraju and Mr. A. Belochitsky. This study is based upon work supported by the Department of Energy Grant No. ER63197-0007185.

References

Chevallier, F., Chéruy, F., Scott, N.A., Chedin, A., 1998. A neural network approach for a fast and accurate computation of longwave radiative budget. J. Appl. Meteorol. 37, 1385–1397.

Chevallier, F., Morcrette, J.-J., Chéruy, F., Scott, N.A., 2000. Use of a neural-network-based longwave radiative transfer scheme in the EMCWF atmospheric model. Quart. J. R. Meteorol. Soc. 126, 761–776.

Chou, M.-D., Suarez, M.J., Liang, X.-Z., Yan, M.M.-H., 2001. A thermal infrared radiation parameterization for atmospheric studies. In: Suarez, Max J. (Ed.), Technical Report Series on Global Modeling and Data Assimilation, NASA/TM-2001-104606, vol. 19.

Collins, W.D., 2001. Parameterization of generalized cloud overlap for radiative calculations in general circulation models. J. Atmos. Sci. 58, 3224–3242.

Duffy, P.B., Govindasamy, B., Iorio, J.P., Milovich, J., Sperber, K.R., Taylor, K.E., Wehner, M.F., Thompson, S.L., 2003. High-resolution simulations of global climate. Part 1. Present climate. Clim. Dynam. 21, 371–390, doi:10.1007/s00382-003-0339-z.

Fox-Rabinovitz, M.S., Takacs, L.L., Govindaraju, R.C., 2002. A variable-resolution stretched-grid general circulation model and data assimilation system with multiple areas of interest: Studying the anomalous regional climate events of 1998. J. Geophys. Res. 107 (D24), 4768, doi:10.1029/2002JD002177.

Hasselmann, S., Hasselmann, K., 1985. Computations and parameterizations of the nonlinear energy transfer in a gravity wave spectrum. Part I. A new method for efficient computations of the exact nonlinear transfer integral. J. Phys. Oceanogr. 15, 1369–1377.

Hasselmann, S., et al., 1985. Computations and parameterizations of the nonlinear energy transfer in a gravity wave spectrum. Part II. Parameterization of the nonlinear transfer for application in wave models. J. Phys. Oceanogr. 15, 1378–1391.

Jolliffe, I.T., 2002. Principal Component Analysis. Springer-Verlag, New York, 502 pp.

Journal of Climate, 1998. J. Clim. 11 (6) (the special issue).

Krasnopolsky, V., Breaker, L.C., Gemmill, W.H., 1995. A neural network as a nonlinear transfer function model for retrieving surface wind speeds from the special sensor microwave imager. J. Geophys. Res. 100, 11033–11045.

Krasnopolsky, V.M., January 1997. A neural network forward model for direct assimilation of SSM/I brightness temperatures into atmospheric models. In Research Activities in Atmospheric and Oceanic Modeling, Report No. 25, WMO/TD-No. 792. CAS/JSC Working Group on Numerical Experimentation, pp. 1.29–1.30.

Krasnopolsky, V.M., Breaker, L.C., Chalikov, D., Rao, D.B., February 1999. Using neural networks to solve the inverse problem of extracting salinity from the UNESCO equation of state for sea water and to speed up the evaluation of this equation. In: Ritchie, H. (Ed.), Research Activities in Atmospheric and Oceanic Modelling, Report No. 28, WMO/TD No. 942, pp. 8.26–8.27.

Krasnopolsky, V.M., Chevallier, F., 2001. Some neural network applications in environmental sciences. ECMWF Technical Memorandum No. 359.

Krasnopolsky, V.M., Chevallier, F., 2003. Some neural network applications in environmental sciences. Part II. Advancing computational efficiency of environmental numerical models. Neural Networks 16, 335–348.

Krasnopolsky, V.M., Schiller, H., 2003. Some neural network applications in environmental sciences. Part I. Forward and inverse problems in satellite remote sensing. Neural Networks 16, 321–334.

Krasnopolsky, V.M., Chalikov, D.V., Tolman, H.L., 2002. A neural network technique to improve computational efficiency of numerical oceanic models. Ocean Model. 4, 363–383.

Krasnopolsky, V.M., Fox-Rabinovitz, M.S., Chalikov, D.V., 2004. Using neural networks for fast and accurate approximation of the long wave radiation parameterization in the NCAR community atmospheric model: evaluation of computational performance and accuracy of approximation. In: Proceedings of the 84th AMS Annual Meeting, 15th Symposium on Global Change and Climate Variations, P1.20, Seattle, Washington, January 11–15, 2004 (CD-ROM).

Krasnopolsky, V.M., Fox-Rabinovitz, M.S., Chalikov, D.V., 2005a. New approach to calculation of atmospheric model physics: accurate and fast neural network emulation of long wave radiation in a climate model. Mon. Wea. Rev.

Krasnopolsky, V.M., Fox-Rabinovitz, M.S., Chou, Ming-Dah, 2005b. Robustness of the NN approach to emulating atmospheric long wave radiation in complex climate models. Proceedings of the International Joint Conference on Neural Networks 2005, Montréal, Que., Canada, July 31–August 4, 2005.

Lorenz, E.N., 1956. Empirical orthogonal functions and statistical weather prediction. Sci. Rep. No. 1, Statistical Forecasting Project. M.I.T., Cambridge, MA, 48 pp.

Ripley, B.D., 1997. Pattern Recognition and Neural Networks. Cambridge University Press, 403 pp.

Schoendorf, J., Rabitz, H., Li, G., 2003. A fast and accurate operational model of ionospheric electron density. Geophys. Res. Lett. 30, 1492–1495.

Stogryn, A.P., Butler, C.T., Bartolac, T.J., 1994. Ocean surface wind retrievals from special sensor microwave imager data with neural networks. J. Geophys. Res. 90, 981–984.

Tolman, H.L., Krasnopolsky, V.M., Chalikov, D.V., 2005. Neural network approximations for nonlinear interactions in wind wave spectra: direct mapping for wind seas in deep water. Ocean Model. 8, 253–278.